Note on Gauss-Newton method

<u>basics</u> Scalar valued function of scalar : $f(x) \Rightarrow f'(x), f''(x)$

Generalizations of derivatives.

Scalar valued function of <u>vector</u> : $f(\bar{x}) \to \underbrace{\nabla f(\bar{x})}_{\substack{\text{vector}\\ \text{gradient}}}, \underbrace{\nabla^2 f(\bar{x})}_{\substack{\text{matrix}\\ \text{hessian}}}$

vector valued function of a vector $\bar{F}(\bar{x}) \Rightarrow$ Jacobian

we have set of points (t_i, y_i) ; $i = 1, \ldots, m$ (m points)

Function $F(\bar{x}, t)$. We want to fit data in non-linear least squares sense. $\bar{x}$ is a vector of model parameters.

$$\text{Define } g(\bar{x}) = \frac{1}{2} \|r(\bar{x})\|_2^2 = \frac{1}{2} \sum_{i=1}^{m} \left(r_i(\bar{x})\right)^2$$

$$r_i(\bar{x}) = \underbrace{y_i}_{\text{scalar}} - \underbrace{F(\bar{x}, t_i)}_{\text{scalar}} = \text{scalar}$$

$r_i(\bar{x})$ is a scalar valued function of a vector.

$$\text{We seek } \bar{x} = \arg\min_{x} \left\{\tfrac{1}{2}\|r(\bar{x})\|^2\right\} = \arg\min_{x} \{g(\bar{x})\}$$

local min found by setting $\nabla g(\bar{x}) = 0$.

This represents a set of non-linear equations.

Newton's Method for $\nabla g(\bar{x})$ yields the scheme:

$$\bar{x}_{n+1} = \bar{x}_n - \left[\nabla^2 g(\bar{x}_n)\right]^{-1} \nabla g(\bar{x}_n)$$

why? consider scalar nonlinear function $f(x)$.

$f(x) = 0$ root finding by NM gives: $x_{n+1} = x_n - \dfrac{f(x_n)}{f'(x_n)}$

$f'(x) = 0$ extreme values finding by NM : $x_{n+1} = x_n - \dfrac{f'(x_n)}{f''(x_n)}$

Replacing $f'(x) = 0$ by $\nabla g(\bar{x}) = 0$ yields above form.

How to get $\nabla g(\bar{x})$ and $\nabla^2 g(\bar{x})$?

$$r(\bar{x}) = \begin{pmatrix} r_1(\bar{x}) \\ \vdots \\ r_m(\bar{x}) \end{pmatrix}$$

note that $r(\bar{x})$ is a vector valued function of a vector.
→ generalization of the derivative for $r(\bar{x})$ is the Jacobian.

recall that $r(\bar{x}) = \bar{y} - F(\bar{x}, \bar{t})$

$$r_i(\bar{x}) = y_i - F(\bar{x}, t_i)$$

$$J[r(\bar{x})]_{ij} = \frac{\partial r_i(\bar{x})}{\partial x_j} = \frac{-\partial F(\bar{x}, t_i)}{\partial x_j}$$

$$\nabla r_i(\bar{x}) = \left[\frac{\partial r_i}{\partial x_1}, \ldots, \frac{\partial r_i}{\partial x_m} \right]^T$$ gradient of scalar function $r_i(\bar{x})$

It thus follows that, since y_i is a scalar:

$$J[r(\bar{x})]_{i,:} = \nabla r_i(\bar{x})^T = -\nabla F(\bar{x}, t_i)^T$$

We must take transpose to set the i-th row of the Jacobian matrix to a row vector.

Now $g(\bar{x}) = \frac{1}{2} \|r(\bar{x})\|^2 = \frac{1}{2} \sum_{i=1}^{m} r_i(\bar{x})^2$

$$\Rightarrow \nabla g(\bar{x}) = \sum_{i=1}^{m} \underbrace{r_i(\bar{x})}_{\text{scalar}} \underbrace{\nabla r_i(\bar{x})}_{\text{vector}} = \underbrace{J[r(\bar{x})]^T}_{\text{matrix}} \underbrace{r(\bar{x})}_{\text{vector}}$$

It remains to find the Hessian $\nabla^2 g(\bar{x})$.

$$\nabla^2 g(\bar{x}) = \sum_{i=1}^{m} \underbrace{\nabla r_i(\bar{x}) \nabla r_i(\bar{x})^T}_{\text{matrix}} + \sum_{i=1}^{m} \underbrace{r_i(\bar{x}) \nabla^2 r_i(\bar{x})}_{\text{scalar} \times \text{matrix}}$$

This follows by the vv^T gradient "product rule".

Using the relationship between Jacobian and gradient, we get:

$$\nabla^2 g(\bar{x}) = J^T J + \sum_{i=1}^{m} r_i(\bar{x}) \nabla^2 r_i(\bar{x})$$

where $J = J[r(\bar{x})]$ Jacobian of vector valued residual function

For Newton-Gauss method, we use the approximation:

$\nabla^2 g(\bar{x}) \simeq J^T J$ dropping the higher order terms

Notice $J[r(\bar{x})]$ and $\nabla r_i(\bar{x})$ are expressible in terms of $J[F(\bar{x}, \bar{t})]$ and $\nabla F(\bar{x}, t_i)$.

Thus, NG method yields:

$\bar{x}_{n+1} = \bar{x}_n - \alpha_n [J_n^T J_n]^{-1} J_n^T r_n$ with $0 < \alpha_n \leq 1$ a step size parameter determined via line search.

$\underbrace{[J^T J]\, y = J^T r}_{\text{linear system solve.}} \Rightarrow y = [J^T J]^{-1} J^T r$

Number of ways to solve the linear system: either CG type schemes (since $J^T J$ is SPD) or direct solve.

Ex) Let $J^T J = M$, $J^T r = b \Rightarrow My = b$

Let $PM = LU \Rightarrow PMy = Pb \Rightarrow LUy = Pb$

$\Rightarrow z = Uy \Rightarrow Lz = Pb \Rightarrow$ then solve $Uy = z$ for y.